Establishing Context in Task-Oriented Dialogs

Barbara G. Deutsch

Artificial Intelligence Center
Stanford Research Institute
Menlo Park, California 94025

ABSTRACT

This paper describes part of the discourse component of a speech understanding system for task-oriented dialogs, specifically, a mechanism for establishing a focus of attention to aid in identifying the referents of definite noun phrases. In building a representation of the dialog context, the discourse processor takes advantage of the fact that task-oriented dialogs have a structure that closely parallels the structure of the task. The semantic network of the system is partitioned into focus spaces with each focus space containing only those concepts pertinent to the dialog relating to a subtask. The focus spaces are linked to their corresponding subtasks and ordered in a hierarchy determined by the relations among subtasks.

INTRODUCTION

Language communication entails the transmission of concepts from the speaker's model of the world to the listener's. It is crucial that the speaker be able to communicate descriptions of concepts in his model in a way that allows the listener to pick out the relevant related concept in his model. In normal human communication it is not necessary to describe a concept in a completely unambiguous way. Contextual clues from both the situation and the surrounding dialog are counted on to help disambiguate. The listener's problem is to use that context to help in his identification of the concept being communicated. As a simple example, consider the utterance, "Hand me the box-end wrench," as it might occur in a conversation between two people working on a maintenance task. Although many box-end wrenches may be known to both the speaker and the listener, the fact that the listener has a particular box-end wrench in his hand makes the noun phrase unambiguous. (For other examples, see Norman, Rumelhart, et al., 1975). In the most extreme case, the use of pronouns depends entirely on the dialog context to determine the intended referent; "it" can refer to any single inanimate object or event.

A related problem arises with elliptical expressions. Often the surrounding dialog supplies enough information so that only a word or two suffices to communicate an entire (complex) idea. For example, consider the following exchange:

E: Bolt the pump to the platform.
A: O.K.
E: What tools are you using [to bolt the pump
   to the platform].
A: My fingers [are the tools I am using ...]

The expressions in brackets indicate the full utterance that was meant by the partial utterance. The listener must fill in this information from the surrounding dialog.

This paper considers such phenomena as they occur in task-oriented dialogs. By task-oriented dialog we mean conversation directed toward the completion of some task. In particular, we will be concerned with a computer-based consultant task in which an apprentice technician communicates with a computer system about the repair of electromechanical devices. The understanding system must maintain models of the world and of the dialog to disambiguate references in the apprentice's speech.

DISCOURSE IN SPEECH UNDERSTANDING

In a speech understanding system, the discourse component is one of several sources of knowledge that must interact in interpreting an utterance (see Paxton and A. Robinson, 1975; J. Robinson, 1975). Because of the uncertainty in the acoustic signal, it is important that higher level sources of knowledge like discourse give advice to the system at early stages in the analysis. For this reason, in our current speech system, routines for identifying the referents of definite noun phrases are applied as soon as a possible noun phrase is identified rather than waiting for an interpretation of the entire

utterance. In essence, the procedure entails searching the recent context to find possible referents and returning a list of candidates.

Ellipsis and pronoun resolution require a more local context than the resolution of nonpronominal definite noun phrases (DNPs). A description of the processing for ellipsis and pronoun resolution is contained in the section "Discourse Analysis and Pragmatics" in Walker et al., 1975. In this paper we concentrate on mechanisms for resolving DNPs.

DEFINITE NOUN PHRASES

The problem of resolving DNPs is basically a problem of finding a matching structure in memory. In the case of a computer system with a semantic network knowledge base, the problem is that of finding the network structure corresponding to the structure of the noun phrase. The node that maps onto the head node of the parse structure representing the noun phrase is the concept being identified by the noun phrase. For example, if the knowledge base contains the nodes shown in Figure 1 (and there are no other nodes with e (element) or s (superset) arcs to wrenches), then either node W1 or node W3, but not W2, will match the phrase "the box-end wrench". Matching is not always so straightforward. For example, consider the situation portrayed in Figure 2. The ed, or delineating element, arc (see Hendrix, 1975a) links a node to delineating information about members of the class that node represents. B-E is a set of
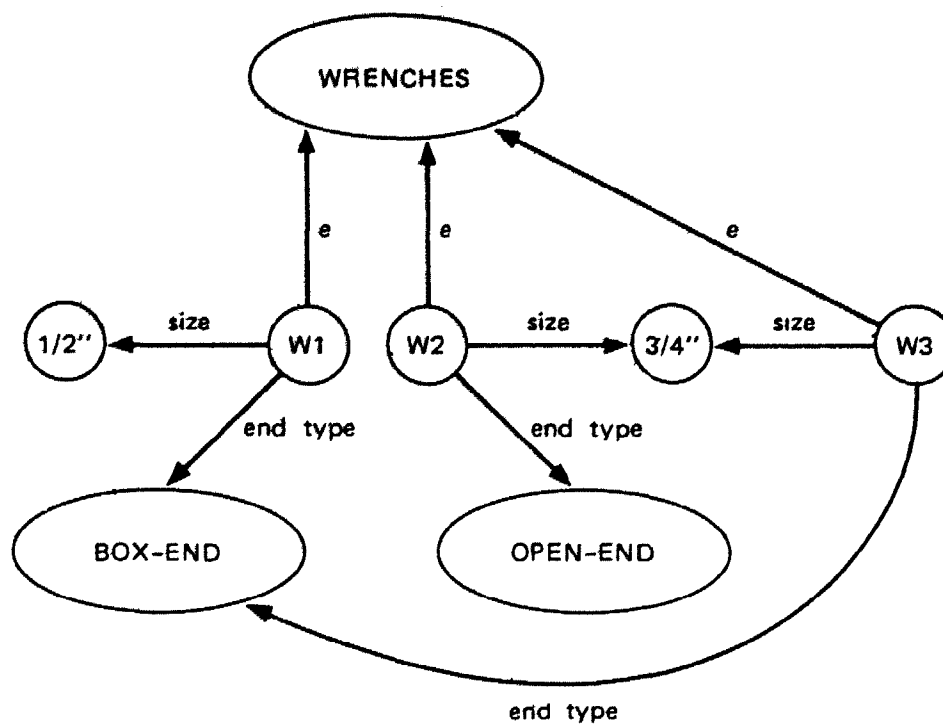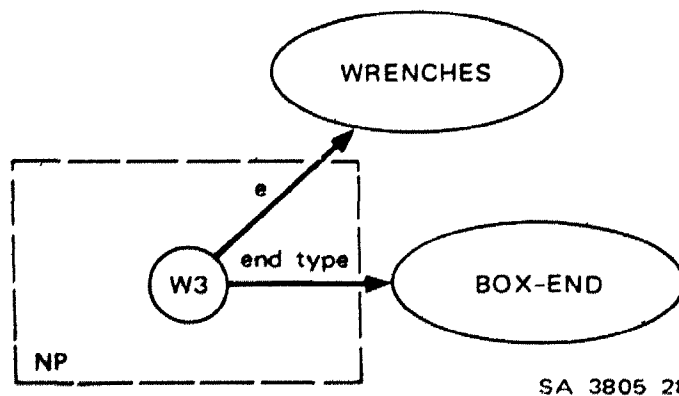
FIGURE 1   NETWORK DESCRIPTION OF THREE WRENCHES
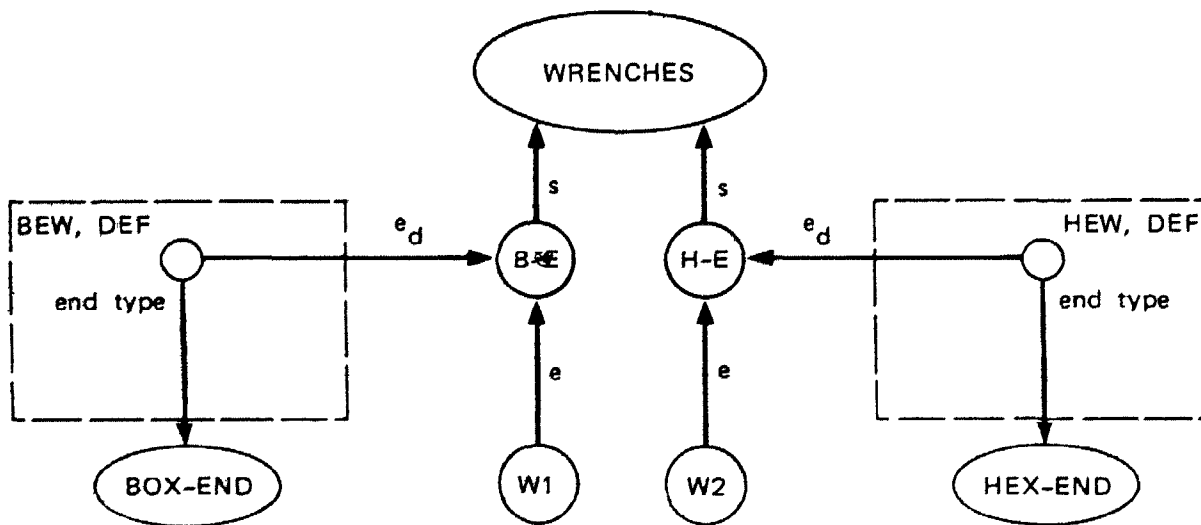
FIGURE 2   SEMANTIC NET SHOWING MEMBERS OF TWO SUBSETS OF THE SET "WRENCHES"

FIGURE 3   SEMANTIC NET SHOWING PARSE SPACE FOR "BOX-END WRENCH"

box-end wrenches to which W1 belongs. H-E is a set of hex-end wrenches to which W2 belongs. If the apprentice now says, "... the box-end wrench", he means W1. The utterance level structure created by parsing (see Hendrix, 1975b) for the phrase "the box-end wrench" is inside the space NP in Figure 3; some deduction must be done to establish the correspondence between W1 and W3.

The structure matching routines that form a basic part of the DNP resolver take as inputs a parse level network of nodes and arcs and a data network to match it against. (The current matcher was written by R. E. Fikes). In general, a large number of objects in the data net may be candidates for the matcher (i.e., objects that are elements of the same set as the object being identified by the DNP). Since, in itself, the matcher has no way of deciding which objects to consider first, additional mechanisms are needed to limit the search.

FOCUS SPACES

The discourse component must determine a subnet of the semantic net knowledge base for consideration by the matcher. That is, it must be able to establish as a local context that subset of the system's total knowledge base that is relevant at a given point in the dialog. This is analogous to determining what is in the user's focus of attention. Put another way, we would like to highlight certain nodes and arcs of the semantic network.

In task-oriented dialogs, the dialog context is actually a

composite of three different component contexts: a verbal context, a task context, and a context of general world knowledge. The verbal context includes the history of preceding utterances, their syntactic form, the objects and actions discussed in them, and the particular words used. The task context is the focus supplied by the task being worked on. It includes such information as: where the current subtask fits in the overall plan, what its subtasks are, what actions are likely to follow, what objects are important. The context of general world knowledge is the information that reflects a background understanding of the properties and interrelations of objects and actions: for example, the fact that tool boxes typically contain tools and that attaching entails some kind of fastening.

To highlight objects in the dialog and provide verbal context, network partitioning is used in a new way. Hendrix (1975a) has suggested imposing a logical partitioning on network structures for encoding logical connectives and quantifiers. Using the same technique, a focus partitioning may be used to divide the network 'nto a number of local contexts. Nodes and arcs belong to both logical and focus spaces. The logical and focus partitions are independent of one another in the sense that the logical spaces on which a node or arc lies neither determine nor depend on the focus spaces in which the node or arc lies.

A new focus space is created for each subtask that enters the dialog. The task model (described shortly) imposes a hierarchical ordering, based on the subtask hierarchy, on these

spaces. This hierarchy determines what nodes and arcs are visible from a given space. The arcs and nodes that belong to a space are the only ones immediately visible from that space. Arcs and nodes in spaces that are above a given space in the hierarchy are potentially visible, but must be requested specifically to be seen. Other arcs and nodes are not visible.

A node may appear in any number of focus spaces. When the same object is used in two different subtasks, either the same or different aspects of the object may be in focus in the two subtasks. It is also possible for a node or arc to be in no focus space. In this case, the object is not strongly associated with the actual performance of any particular subtask. Such objects must be described relative to the global task environment. For completeness, we define a top-most space, called the "communal space", and a bottom-most space, called the "vista space". The communal space contains the relationships that are time invariant (e.g., the fact that tools are found in tool boxes) or common to all contexts. The vista space is below all other spaces and hence can see everything in the semantic net. This perspective is useful for determining all the relationships into which an object has entered.

The task model in our system will be embodied in a procedural net which encodes the task structure in a hierarchy of subtasks and encodes each subtask as a partial ordering of steps (Sacerdoti, 1975). The procedural net system also allows tasks to be expanded dynamically to further levels of detail when

necessary. A representation of the hierarchy of subtasks is important for reference resolution. An examination of task-oriented dialogs shows that references operate within tasks and up the hierarchy chain (Deutsch, 1974). Using the hierarchy of the procedural net to impose a hierarchy on the focus spaces enables us to search for references in hierarchical order. Having a representation of the partial ordering of tasks allows us to capture the alternatives the apprentice has in choosing subsequent tasks.

We have explicitly separated the three components of the dialog context. The representation of an object in a focus space will include only the relationships that have been mentioned in the dialog concerning the corresponding subtask or that are inherent in the procedural net description of the local task. Thus, the verbal component is supplied by the information recorded in the focus space hierarchy. Forward references to objects in the task (task component) are found by examining the procedural net. The general world knowledge component is information that is present in the communal space. When resolving a DNP, we can dynamically allocate effort between examining links in the local focus space, looking forward in the task, looking back up the focus space hierarchy, and looking deeper into knowledge base information.

GENERAL STRATEGY

The strategy we are currently exploring is first to examine

the currently active focus space and then to examine the next level of detail in the task. If the referent cannot be found in either of these locations, we look up the focus space hierarchy and then further down the task chain. The current context to be used by the discourse processor includes:

(1) A focus space containing the objects currently in focus
(2) A link to the associated node in the task model
(3) A type flag used in setting up expectations.

The type is necessary because there are subdialogs that do not directly reflect on the task structure. For example, there are subdialogs about tool identification ("What is a wheelpuller?") and tool use ("How do I use this wrench?"). References in these subdialogs do not follow the same focus space hierarchy and task structure.

The dialog shown in Table 1 will be examined to show how a combination of a task model and focus spaces may be used to help resolve DNPs.

E: I would like you to assemble the air compressor.
A: O.K.
E: I suggest you begin by attaching the pump to the platform.
A: O.K.
E: What are you doing now?
A: Using the pliers to get the nuts in underneath the platform.
E: I realize this is a difficult task.
A: I'm tightening the bolts now. They're all in place.
E: Good.
A: How tightly should I install this pipe elbow that fits into the pump?

Table 1: Subdialog for aircompressor assembly.

A partial procedural net for assembling an air compressor is

shown in Figure 4. The terms "install", "connect", "attach" refer to conceptual actions rather than lexical items. The dashed lines connect higher level tasks to their constituent subtasks. The time sequence of steps in the task is left to right. The partial ordering of tasks is encoded with the S and J nodes. The S, or ANDSPLIT, node indicates the beginning of parallel branches in the partial ordering. The nodes on arcs coming out of an S node may be done in any order. The J, or ANDJOIN, node indicates a point where several parallel tasks must be completed. The boxes labeled T are relevant to the subdialog fragment.

In the following analysis of the dialog, the utterances are considered sequentially. DNP resolution is considered in relation to the dialog history and the procedural net task model. (The search for references inside focus spaces is currently implemented; integration with the task model is not.) The context information listed under (1)-(3) above is shown in the accompanying figures as follows: (1) label on spaces in the network; (2) PNETTIE; (3) FSTYPE.

E: I would like you to assemble the air compressor.
A: O.K.
E: I suggest you begin by attaching the pump to the platform.

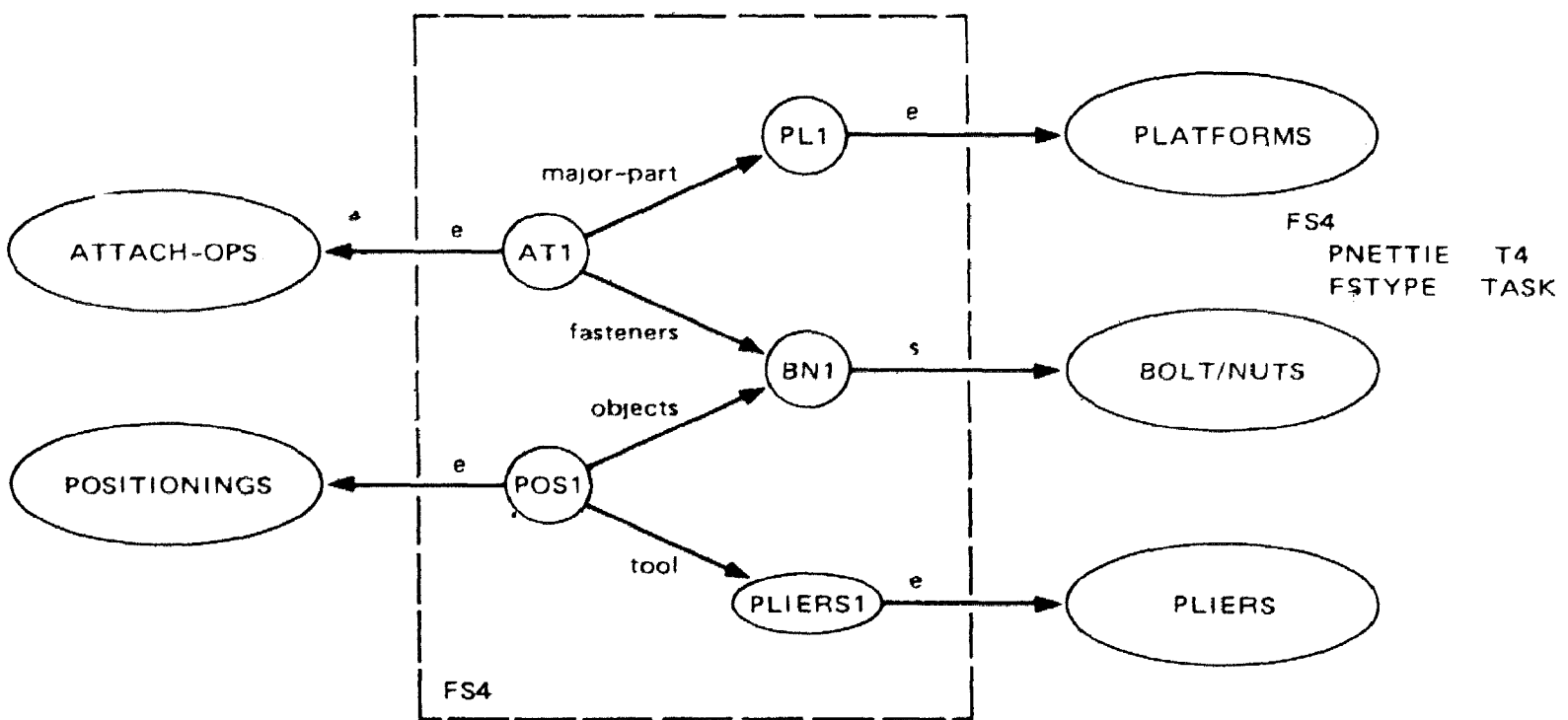[At this point, we are at task T1; focus spaces FS0 and FS1 shown in Figure 5 have been set up.]

A: O.K.

[This could mean I'm done, but the response comes right after the instruction and the task takes a while.]

TA-740522-84

FIGURE 4   PARTIAL PROCEDURAL NET FOR ASSEMBLING AIR COMPRESSOR
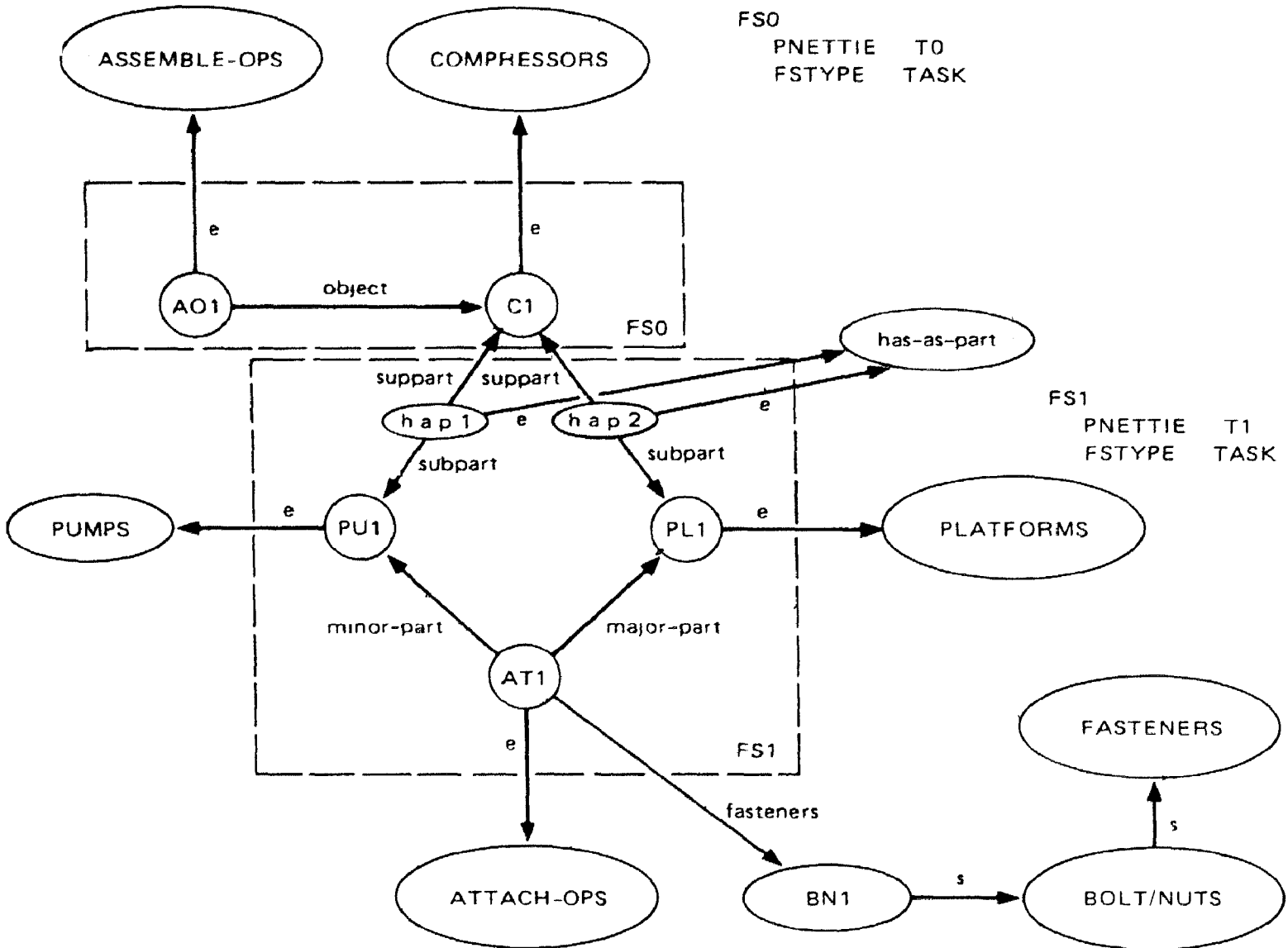
FIGURE 5   FOCUS SPACES FSO AND FS1

FIGURE 6   FOCUS SPACE FOR STARTING BOLT/NUTS OPERATION

E: What are you doing now?

[After a suitable waiting period, the expert queries the progress of the user.]

A: Using the pliers to get the nuts in underneath the platform.

["The pliers" can be resolved because there is only one pair; if this were not the case, the task model would have to be consulted. For both "the nuts" and "the platform", the FS hierarchy is consulted. "The platform", P1 is in focus in the current FS. There is no sign of nuts so we look forward in the task model. The relevant parts are located in subtask T4. This causes a new context, to be established as shown in Figure 6.]

E: I realize this is a difficult task.

[An attempt to assess the apprentice's perception of the problem. Note that at this point the task has barely begun and the expert does not have a very good model of the apprentice.]

A: I'm tightening the bolts now. They're all in place.

[FS4 contains "the bolts"; they were brought into focus when T4 was started. "They" is determined to refer to "the bolts" by checking the objects in the previous utterance for number agreement. Note that the last statement confirms the closure of T4. "Tighten" opens T5.]

E: Good.
A: How tightly should I install this pipe elbow that fits into the pump?

[There is no pipe elbow in the current FS. (Note that up until that point in the query the apprentice might have been asking about task T5). We close T5; because of the task structure this brings us back up to the top level. We are at the point of looking into new tasks. At present all of the tasks are considered equally. Eventually T6 is found to involve an elbow.]

In summation, then, the focus spaces provide a way of isolating certain parts of the semantic net, thus providing a way to focus on immediately relevant information. By tying the focus spaces to a model of the task, we are able to consider forward

task references. Both the task model and the focus spaces are linked to the general knowledge base; thus, it is possible to go from an item in either the task model or a focus space to other known but not previously referenced information about that item. The focus spaces and task model provide access to context information about objects in the domain, making it possible to focus on a relevant subset of the system's knowledge.

References

Deutsch, Barbara G. The Structure of Task-Oriented Dialogs. Contributed Papers, IEEE Symposium on Speech Recognition, Carnegie-Mellon University, Pittsburgh, Pennsylvania, 15-19 April 1974. IEEE, New York. 1974, 250-254.

Hendrix, Gary G. Expanding the Utility of Semantic Networks Through Partitioning. Advance Papers of the Fourth International Joint Conference on Artificial Intelligence, Tbilisi, Georgia, USSR, 3-8 September 1975, 115-121 (a).

Hendrix, Gary G. Semantic Processing for Speech Understanding. Presented at the Thirteenth Annual Meeting of the Association for Computational Linguistics, Boston, Massachusetts, 30 October - 1 November 1975 (b).

Norman, D. A., Rumelhart, D. E., et al., Explorations in Cognition. W. H. Freeman and Company, San Francisco, 1975.

Paxton, William H., and Robinson, Ann E. System Integration and Control in a Speech Understanding System. Presented at the Thirteenth Annual Meeting of the Association for Computational Linguistics, Boston, Massachusetts, 30 October - 1 November 1975.

Robinson, Jane J. A Tuneable Performance Grammar. Presented at the Thirteenth Annual Meeting of the Association for Computational Linguistics, Boston, Massachusetts, 30 October - 1 November 1975.

Sacerdoti, Earl. A Structure for Plans and Behavior. Technical Note 109, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California, August 1975.

Walker, Donald E., et al. Speech Understanding Research. Annual Report, Project 3804, Artificial Intelligence Center, Stanford Research Institute, Menlo Park, California, June 1975.